

Sachliche Erschließung und die neue Suchmaschine des Österreichischen Bibliothekenverbundes

Markus Knitel, Josef Labner und Otto Oberhauser

Mit der Einführung der neuen Suchmaschine Primo im Österreichischen Bibliothekenverbund wurde eine Neubewertung und Neuausrichtung der Anzeige, Indexierung und der erstmals möglichen Facettierung sacherschließender Elemente notwendig. Nach einem kurzen Überblick über die im Verbund zum Einsatz kommende Software sowie über Umfang und Art der inhaltlichen Erschließung im zentralen Katalog wird detailliert auf die angestrebte Umsetzung derselben in der neuen Suchmaschinenumgebung eingegangen, wobei der verbalen Erschließung durch RSWK-Ketten als wichtigste Form der intellektuellen Erschließung erhöhte Aufmerksamkeit zukommt. In einem zweiten Teil wird die konkrete technische Umsetzung in Primo beleuchtet, um abschließend einen Ausblick auf zukünftige Entwicklungen sowie projektierte Verbesserungen zu geben.

Subject indexing and the new search engine of the Austrian Library Network

Recently, the Austrian Library Network implemented its new search engine, Primo. Hence, a reappraisal and realignment was needed for the display, indexing and – for the first time – faceted navigation of all data elements relevant for subject access. After a short overview of the software and the status of subject indexing and classification in the union catalogue of the consortium, the article deals in detail with the intended implementation in the new search engine. The main emphasis is on subject headings (according to the German subject cataloguing rules) as the most important means of intellectual indexing. In the second part, the actual implementation in Primo is outlined. Finally, future developments and improvements are discussed.

Sujet d'indexation et le nouveau moteur de recherche du réseau autrichien de bibliothèques

L'article traite la configuration du catalogage matières pour le nouveau moteur de recherche Primo du réseau autrichien de bibliothèques. Après un court tour d'horizon du réseau autrichien et du logiciel utilisé suit une analyse de l'utilisation des différentes formes du catalogage matières, c'est à dire des mots-sujet, des classifications et des résumés analytiques dans le catalogue central. Ensuite les auteurs présentent leurs idées pour l'emploi le mieux possible de ces éléments dans l'affichage de notices complètes et courtes, l'indexation et les facettes pour ensuite expliquer la réalisation technique dans Primo. Finalement l'article s'occupe de développements et de perfectionnements futures.

1 Einleitung

Der Österreichische Bibliothekenverbund ist der nationale Verbund universitärer, wissenschaftlicher und administrativer Bibliotheken des Landes. Über 75 Bibliotheken von ca. 60 Trägerinstitutionen nehmen aktiv durch Online-Katalogisierung an diesem Verbund teil, darunter die Österreichische Nationalbibliothek sowie nahezu alle bundesstaatlichen Universitätsbibliotheken. Von weiteren 310 Einrichtungen werden die Zeitschriftenbestände sowie Buchbestände aus einem früheren Gesamtkatalog nachgewiesen. Seit Anfang 1999 wird verbundweit das integrierte Bibliothekssystem Aleph 500 des Herstellers Ex Libris eingesetzt. Unter diesem System werden sowohl der Österreichische Verbundkatalog,¹ der

von der Verbundzentrale OBVSG² betriebene Gesamtkatalog, als auch die lokalen Online-Kataloge der einzelnen Verbundbibliotheken – einige davon auch mit dem kleineren System Alephino – frei zugänglich angeboten.

Schwerpunktmäßig umfasst der Verbundkatalog Literatur ab dem Erscheinungsjahr 1980. Von den meisten Verbundteilnehmern wurde und wird auch ältere Literatur durch Nachbearbeitung einzelner Titel oder im Rahmen größerer Retrokatalogisierungsprojekte erfasst. Zunehmend werden auch elektronische Dokumente nachgewiesen. Als Service zur Anreicherung des Verbundkataloges mit letzteren wurde von der OBVSG vor mehreren Jahren das System eDOC eingerichtet, das Objekttypen wie Inhaltsverzeichnisse, Abstracts, Rezensionen, Umschlagbilder und Klappentexte sowie Volltextdokumente unterstützt und mit dem Verbundkatalog automatisch verlinkt. Die Recherche³ in eDOC erfolgt mittels der Suchmaschine Swish-e.

Der Gesamtkatalog des Österreichischen Bibliothekenverbundes weist mit Stand Ende Juni 2010 rund 7,7 Mio. Titel, 14,5 Mio. Exemplare und 0,7 Mio. Zeitschriften-Bestandsangaben nach. In eDOC sind zum selben Zeitpunkt rund elektronische 420.000 Dokumente enthalten.

Das jüngste Produkt in der Softwarepalette des Verbundes ist die neue Retrieval-Software Primo, ebenfalls aus dem Hause Ex Libris. Die Einführung eines neuen Retrievalinstrumentes war lange überfällig. Der herkömmliche Aleph-OPAC, der vom Hersteller schon seit geraumer Zeit nicht mehr weiterentwickelt wurde, kann benutzerseitigen Erwartungen und Gepflogenheiten, die aus dem täglichen Umgang mit bekannten Volltextsuchmaschinen resultieren, nicht adäquat begegnen. Aus Statistiken und Beobachtungen ist bekannt,

1 <http://opac.obvsg.at/acc01> [alle im folgenden zitierten Web-Adressen überprüft am 06.07.2010].

2 Die Österreichische Bibliothekenverbund und Service GmbH (OBVSG) ist ein Unternehmen im Eigentum der Republik Österreich mit Sitz in Wien (<http://www.obvsg.at>).

3 <http://media.obvsg.at/>

```

<record>
+<control></control>
-<display>
  <type>article</type>
  <title>Die Paradoxie des Entscheidens</title>
  <creator>Luhmann, Niklas</creator>
-<ispartof>
  In: Paradoxien der Entscheidung / Friedrich Balke ... (Hg.). - Bielefeld, 2004. - S. 17-
</ispartof>
  <language>ger</language>
  <source>ACC_aleph</source>
  <avallibrary>$$IACC_B_UBW$$Scheck_holdings$$LACC_B_UBW</avallibrary>
  <avallibrary>$$IACC_B_UBI$$Scheck_holdings$$LACC_B_UBI</avallibrary>
  <lds02>AC04113921</lds02>
  <lds30>AC05094114</lds30>
  <lds31>!!!-UBI-!!!</lds31>
  <avallpnx>available</avallpnx>
</display>
+<links></links>
+<search></search>
-<sort>
  <title>DIE PARADOXIE DES ENTSCHEIDENS</title>
  <creationdate>2004</creationdate>
</sort>
+<facets></facets>
+<dedup></dedup>
+<frbr></frbr>
+<delivery></delivery>
+<ranking></ranking>
+<addata></addata>
</record>

```

Abbildung 1

dass die anspruchsvolleren Funktionalitäten (Index blättern, Expertenmodus) nur von einer kleinen Minderheit qualifizierter User genutzt werden, während die große Mehrheit die simple Stichwortrecherche im Suchschlitz der „Alle-Felder-Suche“ bevorzugt. Dass dieses Suchverhalten nur allzu oft zu unbefriedigenden Ergebnissen führt – insbesondere da der OPAC keine sinnvolle Relevanzsortierung der Treffer erlaubt⁴ –, ist ein manifester Nachteil, den es auszugleichen galt. Darüber hinaus war in den letzten Jahren im Bereich der lokalen Bibliotheken, aber auch des Gesamtkataloges (vgl. das oben erwähnte eDoc) eine zunehmende Zersplitterung des Informationsangebotes in unterschiedliche Sucheinstiege, Datenbanken, Teilkataloge oder Repositorien feststellbar. Die Kombination all dieser Angebote in einer einheitlichen, einfachen Oberfläche war folglich eine der Kernanforderungen an einen neuen Sucheinstieg. Neben dem Relevance Ranking war auch die Verfügbarkeit einer facettierten Navigation⁵ ein wichtiges Kriterium im Hinblick auf einen „next-generation library catalog“. Die Entscheidung für Primo fiel nach Prüfung durch eine Arbeitsgruppe der Vollversammlung des Verbundes im Spätherbst 2008. Aus Gründen der Kostenökonomie sowie auf Grund des bereits sehr hohen Integrationsgrades des Österreichischen Bibliothekenverbundes entschied man sich damals für den Betrieb einer ein-

zigen zentralen Primo-Instanz (angesiedelt bei der OBVSG), die durch mehrere Einrichtungen gemeinsam genutzt werden sollte. Trotz der abzusehenden Nachteile bzw. Einschränkungen bei manchen instanzweiten Konfigurationen wurde davon ausgegangen, dass die Synergieeffekte alle Nachteile überwiegen würden. Inzwischen sind die Kataloge des Verbundes sowie der Universitätsbibliotheken Wien und Innsbruck mit Primo produktiv;⁷ weitere lokale Bibliotheken werden in Kürze folgen.⁸

2 Primo

Herzstück von Primo und Basis der Indexierung ist das *Primo Normalized XML* (PNX), ein programminternes, einfach lesbares XML-Format, das in Sektionen („sections“) gegliedert ist. In dieses Format werden alle Quelldatensätze auf der Basis bestimmter Normalisierungsregeln transformiert. Dabei werden die MAB-Kategorien bestimmten Tags im PNX zugeordnet, welche ihrerseits wiederum für verschiedene Verarbeitungsschritte herangezogen werden: Während die *display section* anzeigerelevante Informationen enthält (z.B. Autor, Titel oder ISBN), befinden sich in der *search section*

die zu indexierenden Anteile des Datensatzes und in der *link section* Links zu Volltexten oder auch zur OPAC-Anzeige.⁹ Als Beispiel zeigt *Abbildung 1* einen PNX-Satz mit den aufgeklappten Sektionen „display“ und „sort“.

Die Quelldaten, in der Regel MAB-XML Datensätze,¹⁰ können für jede Sektion unterschiedlich aufbereitet werden. Das heißt, dass z. B. die Darstellung, die Suchbarkeit und die Facettierung getrennt behandelt werden können. Dass diese Aussage eigentlich sogleich wieder einer Einschränkung bedürfte, wird weiter unten noch deutlich werden. Entgegen gelegentlich geäußelter Befürchtungen macht eine Suchmaschine wie Primo kontrollierte Metadaten keineswegs obsolet – vielmehr kommt die tiefgehende und mühevollere Erschließungsarbeit der letzten Jahre jetzt erst richtig zur Geltung. Auch die in der bisherigen Katalogisierungspraxis ein wenig unterschätzten „festen Felder“ (z.B. MAB 050 = Datenträger, 051 = Veröffentlichungsspezifische Angaben zu begrenzten Werken, 052 = Veröffentlichungsspezifische Angaben zu fortlaufenden Sammelwerken, ...) gewinnen in der neuen Umgebung an Bedeutung.

Die Suchoberfläche selbst sollte bei alledem so schlicht wie möglich gehalten werden. Letztlich obliegt es aber jeder an der zentralen Primo-Instanz teilnehmenden Institution, sich um ihren Internetauftritt zu bemühen. Die Gestaltungsmöglichkeiten sind vielfältig, wenn auch teilweise mit viel Aufwand verbunden. *Abbildung 2* zeigt als Beispiel die Kurzanzeige mit integrierter Suchmaske in der Version des Verbundes. Am frappierendsten und für den Benutzer wohl am offensichtlichsten ist hier einerseits die Sortierung der Treffer nach Relevanz und andererseits das Gruppieren der gefundenen Titel nach unterschiedlichen Facetten, die ein gezieltes Weiter Navigieren und sukzessives Einschränken der Treffermenge in die Richtung des gewünschten Zielles ermöglichen. „Relevanz“ ist das voreingestellte Sortierkriterium; alternativ dazu sind „Erscheinungsjahr“ und „Bekanntheitsgrad“ auswählbar. Mit letzterer Option wird erstmals das Benutzerverhalten in die Treffer-Sortierung einbezogen. Ein Titel kann auf drei Arten an Bekanntheit („popularity“) gewinnen: durch Betrachten in der Vollanzeige, Hinzufügen zur Merklste oder – im Falle der lokalen Bibliotheken – durch

4 Vgl.: Oberhauser, O. (2010). Relevance Ranking in den Online-Katalogen der „nächsten Generation“. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*. 63 (1/2). 25–35.

5 Vgl. z.B.: Imhof, A. (2006). RSWK/SWD und Faceted Browsing: Neue Möglichkeiten einer inhaltlich-intuitiven Navigation. *Bibliotheksdienst*. 40 (8/9). 1015–1025; Olson, T. A. (2007). Utility of a faceted catalog for scholarly research. *Library hi tech*. 25 (4). 550–561.

6 Vgl.: Breeding, M. (2007). Next-generation library catalogs. *Library technology reports*. 43 (4).

7 <http://search.obvsg.at/OBV> (Verbund); <http://usearch.univie.ac.at/> (UB Wien); <http://search.obvsg.at/UB> (UB Innsbruck).

8 Zu Primo im OBV vgl. folgende aktuelle Beiträge: Knitel, M. (2010). Architektur und Parametrisierung von Primo im Österreichischen Bibliothekenverbund: Ein Überblick. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare*. 63 (1/2). 56–68; Babitschev, V.; Brandauer, J. (2010). Die konsortiale Primo-Implementierung im OBV. Teil 1: Aktueller Stand, Herausforderungen, Lösungen. Vortrag, DACHELA 2010, Wien, 15.–16. Juni 2010. Online: http://www.obvsg.at/fileadmin/files/obvsg/publ/primo-dachela-2010_fin.pdf; Krabo, U. (2010). Die konsortiale Primo-Implementierung im OBV. Teil 2: Integration zusätzlicher Dienste. Vortrag, DACHELA 2010, Wien, 15.–16. Juni 2010. Online: http://www.obvsg.at/fileadmin/files/obvsg/publ/primo-dachela-2010_teil2_v1.pdf

9 Primo ist derzeit mittels einer einfachen Verlinkung mit dem jeweils korrespondierenden Aleph-OPAC verbunden. Die im Bereich der lokalen Bibliotheken wichtigen Funktionen wie Bestellung, Einsicht des Benutzerkontos oder Verlängerung ausgeliehener Bücher können bislang nur im herkömmlichen OPAC abgewickelt werden. Dies wird sich mit der neuen Version 3 von Primo ändern.

10 <http://www.d-nb.de/standardisierung/formate/mabxml.htm>

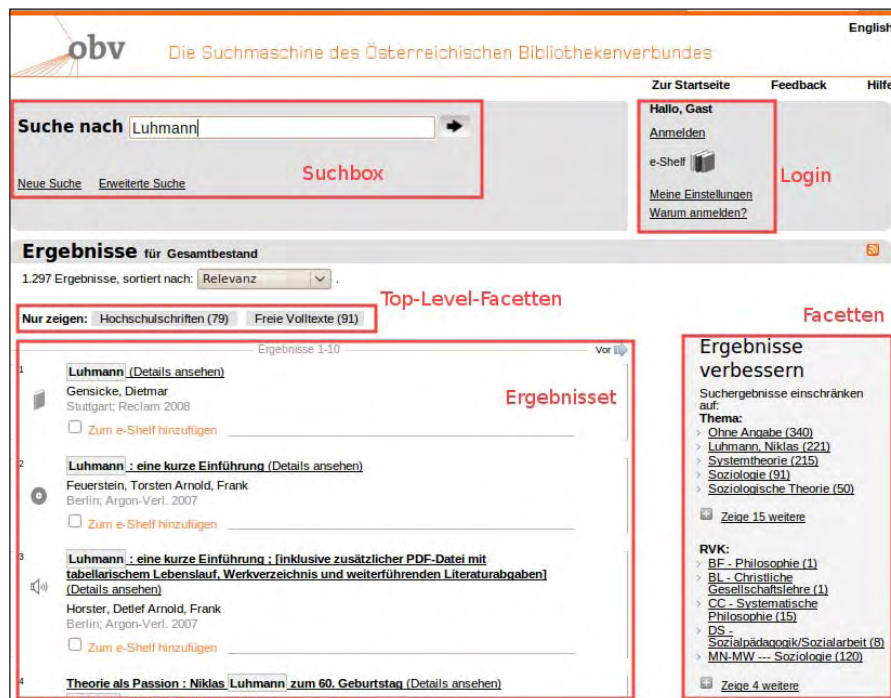


Abbildung 2

Anklicken des sogenannten „getit“-Links, der zur Anzeige der Bestandsinformation in den lokalen OPACs führt. In der Abbildung sind zwei thematische Facetten dargestellt – „Thema“ (RSWK-Schlagwörter) und „RVK“. Welche Facetten bei der Trefferliste aufscheinen (und in welcher Reihenfolge), kann jede teilnehmende Bibliothek selbst bestimmen. In der Version des Verbundes werden neben den beiden genannten auch die Facetten „Medium“, „Form“, „Erscheinungsjahr“, „Sprache“, „Bibliothek“, „Autor/in“ und „Region“ angeboten („Medium“ kann grob als Facette für den Materialtyp wie Buch, CD, Artikel etc. betrachtet werden, während die Facette „Form“ eher nach Kriterien wie Aufsatzsammlung, Kunstführer, Ratgeber etc. gliedert).

Einem bewährten Prinzip folgend, stehen dem Suchenden grundsätzlich eine einfache und eine erweiterte Suche zur Verfügung. Die *einfache Suche* führt eine Recherche im Index „Alle Felder“ durch, während die *erweiterte Suche* eine Beschränkung der Recherche auf bestimmte Indizes, Jahre oder Materialarten erlaubt.

Die beiden eben genannten Hauptfunktionen von Primo und das völlig neue Datenmodell machten es notwendig, den Einsatz und die Präsentation von sacherschließenden Elementen bei der Recherche neu zu überdenken. Was möchte man sichtbar und suchbar, nur sichtbar, oder gar nur suchbar jedoch nicht sichtbar machen? Welche thematischen Facetten sollen angeboten werden? Lohnt sich der Einsatz von Klassifikationen (wenn ja, welcher)? Wie gewichtet man Sacherschließungselemente in der Berechnung der Relevanz? Basis und Aus-

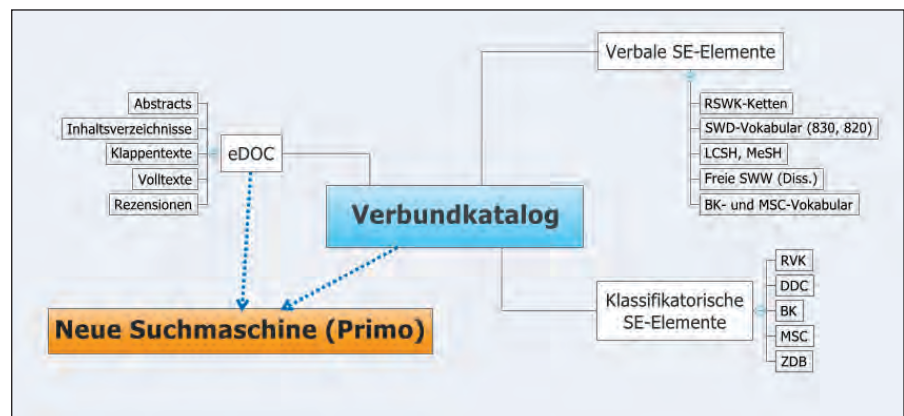


Abbildung 3

gangspunkt der Beantwortung dieser Fragen war die sorgfältige Analyse des Status Quo des Einsatzes von sacherschließenden Elementen. Wir beziehen uns im Folgenden immer nur auf die Ebene des Gesamtverbundes.

3 Sachliche Erschließung im Österreichischen Verbundkatalog

Das gegenwärtig bestehende Erscheinungsbild der sachlichen Erschließung im Österreichischen Bibliothekenverbund ist historisch gewachsen und gleicht – wie in anderen Verbünden auch – ein wenig einem Patchwork-Muster. Es gibt nämlich eine ganze Reihe unterschiedlicher Sacherschließungselemente innerhalb und außerhalb des bisherigen Verbundkatalogs, wie in *Abbildung 3* veranschaulicht wird.

Die **verbale** Sacherschließung, an der sich der überwiegende Teil der Verbundbiblio-

theken beteiligt, erfolgt seit 1989 – kurz nach der Einführung des ersten automatisationsunterstützten Verbundsystems – kooperativ nach RSWK und SWD. Heute weisen 38,3% der Titeldatensätze mindestens *eine* Schlagwortkette nach RSWK bzw. SWD auf. Dieser Anteil betrug 2005 noch 44%,¹¹ ist aber seitdem aufgrund mehrerer großer Retrokonversionsprojekte auf den derzeitigen Stand abgesunken. *Nur eine* Kette weisen 5,3% auf, mindestens 2 Ketten 33%, 3 oder mehr Ketten 19,5%, 4 oder mehr Ketten 8,7%. Durch die Normdatenverknüpfung wird das Vokabular der Schlagwortketten im Aleph-Katalog mit äquivalenten Bezeichnungen (Synonymen) und Alternativen aus der SWD weiter angereichert. Die sachliche Erschließung nach RSWK / SWD spielt unter allen Sacherschließungselementen bei weitem die größte Rolle. Die übrigen *verbale* Sacherschließungselemente, die in den Verbunddaten enthalten sind, sind in quantitativer Hinsicht weniger bedeutend. Der Anteil der im Zuge der

Fremddatennutzung (v.a. BNB) übernommenen Library of Congress Subject Headings (LCSH) beträgt nur 1,8%; jener der auf demselben Weg übernommenen Medical Subject Headings (MeSH) liegt noch weit darunter. Beide Datenelemente gelangen jedoch im Aleph-Katalog nicht nur in den Index „Alle-Felder“, sondern auch in einen kombinierten Schlagwort-Titeltstichwort-Index, in den auch die im Rahmen der Erfassung von österreichischen Dissertationen entstehenden Autoren-Keywords (deutsch und englisch, jeweils deutlich unter 1%) mit eingehen. Schließlich sind noch die Klassenbenennungen aus Basisklassifikation (BK) und Mathematics Subject Classification (MSC) zu erwähnen, mit denen die Titeldaten im Zuge der Normdatenverknüpfung – BK und MSC werden als

¹¹ Vgl.: Oberhauser, O. (2005). Klassifikatorische Erschließung und Recherche im Österreichischen Verbundkatalog. B.I.T. online. 8(2). 123–132.

Normdatei geführt¹² – angereichert werden. Im Bereich der **klassifikatorischen** Sacherschließung gibt es keine verbindliche Verbundklassifikation, sondern fünf Systeme, die in den Verbunddatensätzen zugelassen sind. Von diesen spielt die Regensburger Verbundklassifikation (RVK), die auch in Österreich immer mehr aktive Anwender findet, mit ca. 7,5% (575.000 Titel) die bedeutendste Rolle. Notationen der Dewey Decimal Classification (DDC), die mehrheitlich durch die Nutzung der DNB-Fremddaten in den Verbund gelangen (aktive Anwenderbibliotheken der DDC sind nur die Oberösterreichische Landesbibliothek und Teile der UB Wien), finden sich bei 3,7% der Datensätze. Die BK, die über ein beträchtliches Potential als Grundlage einer sachlichen Facette in einem neuen OPAC verfügen würde,¹³ wird erst seit kurzem von zwei großen Bibliotheken aktiv eingesetzt (Österreichische Nationalbibliothek; UB Wien). Derzeit tragen nur ca. 168.000 Verbundtitel (2,2%) BK-Notationen; eine Datenübernahme (aus dem GBV), mit der es möglich wäre, einen weit höheren Anteil von Titeln mit BK-Notationen anzureichern, ist bisher leider nicht zustande gekommen. Zu erwähnen ist schließlich auch, dass nahezu alle Zeitschriftentitel des Katalogs (ca. 271.000) mit Notationen der ZDB-Systematik versehen sind. Die nur in Fachbereichsbibliotheken eingesetzte MSC ist im Verbundkatalog bei rund 26.000 Titeln vertreten.

In dem als Ergänzung des Verbundkatalogs geführten System **eDOC** werden gegenwärtig rund 420.000 elektronische Objekte mit dem Verbundkatalog verlinkt bzw. indexiert (soweit textuelle Elemente vorhanden sind). Etwa 3,6% der Verbundtitel (ca. 273.000 Datensätze) weisen solche Links auf. Bei 73% der elektronischen Dokumente handelt es sich um Inhaltsverzeichnisse und bei 14,5% um Abstracts.

Der Rest verteilt sich auf Volltexte, Klappentexte, Umschlagbilder, Rezensionen und Titelseiten.

4 Konzeptionelle Überlegungen

Im folgenden werden die Überlegungen zur Rolle der einzelnen sacherschließenden

Datenelemente in der neuen Suchmaschine dargelegt, d.h. die funktionellen Soll-Vorstellungen, wie sie vor der tatsächlichen Implementierung in Primo entstanden sind (nicht immer ist es allerdings dabei möglich, letztere außer Acht zu lassen). Angesichts der quantitativen Gegebenheiten geht es hier hauptsächlich um die RSWK-Schlagwörter; danach wird auch kurz auf andere verbale Sacherschließungselemente sowie die Verbundklassifikationen eingegangen.

4.1 RSWK

Überlegungen zur sachlichen Facette

Wie oben erwähnt, sind im Verbundkatalog die Schlagwortketten nach RSWK das am weitesten verbreitete Mittel der Sacherschließung, auch wenn eine Quote von knapp 40% bei weitem nicht flächendeckend ist. Trotz der von ihm thematisierten Unzulänglichkeiten von RSWK bzw. SWD spricht sich Imhof¹⁴ für die Verwendung der RSWK-Ketten – insbesondere des jeweiligen Hauptschlagwortes – im Rahmen einer facettierten Navigation aus, wobei er für die differenzierte Behandlung der unterschiedlichen Schlagworttypen in verschiedenen Facetten plädiert. Das Problem der auch in Deutschland alles andere als flächendeckenden Erschließung nach RSWK/SWD wird im Rahmen seiner Argumentation freilich außer

SW 1.SW-Kette	902	g	Wien
		a	Stephansdom
		9	4079282-1
SW 1.SW-Kette	902	s	Architektur
		9	4002851-3
SW 1.SW-Kette	902	f	Führer
Erl. 1.SW-Kette	904	a	UBWFRA
SW 2.SW-Kette	907	g	Wien
		a	Stephansdom
		9	4079282-1
SW 2.SW-Kette	907	z	Geschichte
Erl. 2.SW-Kette	909	a	UBWFRA
SW 3.SW-Kette	912	g	Wien
		a	Stephansdom
		9	4079282-1
SW 3.SW-Kette	912	z	Geschichte 1277-1466
SW 3.SW-Kette	912	f	Quelle
Erl. 3.SW-Kette	914	a	UBWFRA

Abbildung 4

Acht gelassen, wenngleich Imhof darauf verweist, dass eine „gute“ Facette idealerweise auf ein Kriterium zurückgreifen können sollte, welches bei 100% der Datensätze anzutreffen ist. Nicht eingegangen wird auch auf die Gefahr, die u.E. bei der Verwendung einer sehr großen Zahl von (spezifischen) Schlagwörtern im Rahmen einer sachlichen Facette besteht: Würde dies nicht dazu führen, dass die Ausprägungen dieser Facette das jeweilige Suchergebnis – anstatt es in handliche Cluster zu untergliedern – in eine Vielzahl von lediglich ganz

schwach besetzten Grüppchen aufsplitten? In Katalogen mit facettierter Navigation kann man bereits jetzt Ähnliches bei der Facette „Autor“ bzw. „Körperschaft“ beobachten. Dennoch erschien vorerst die Verwendung der Schlagwörter für die sachliche Facette die einzige sinnvolle Alternative zu sein, da derzeit viel zu wenige Verbunddatensätze mit klassifikatorischen Sacherschließungselementen, die sich hier vermutlich besser eignen würden, ausgestattet sind.

Extraktion von RSWK-Datenelementen

Welche Elemente der RSWK-Ketten sollten nun für eine sachliche Facette extrahiert werden, welche sind für die künftige Sachrecherche und welche für das Display (Vollanzeige bzw. Weiternavigieren) relevant? (Als Sachrecherche bezeichnen wir hier die im Primo-Suchmodus *erweiterte Suche* mögliche Recherche im Feld „Schlagwort“).

- Für die sachliche Facette wären zweifellos Einzel-Schlagwörter und nicht komplette Ketten zu bevorzugen, da letztere zu einer noch stärkeren Aufsplitterung von Ausprägungen mit geringer Fallzahl führen würde. Gemeint ist allerdings jeweils der Inhalt einer Kategorie 9xx (902, 907, ...) und nicht etwa Stichwörter daraus. Beispiele wären etwa: Hundefutter; Wien / Steinhofkirche; Physikotheologischer Gottesbeweis.
- Die vollständigen RSWK-Ketten werden hingegen für das Display sehr wohl benötigt.
- Form- und Zeitschlagwörter sollen in der Facette unterdrückt werden, für Suche und Vollanzeige hingegen zur Verfügung stehen. Ursprünglich war daran gedacht worden, sie in der Vollanzeige zwar darzustellen, aber auf ihrer Basis kein Weiternavigieren zuzulassen.
- Geographika ohne Unterfeld \$9, d.h. ohne Normdatenverlinkung, wären in der Facette ebenfalls zu unterdrücken (dies betrifft in erster Linie Kongress-/Ausstellungsorte), sollen aber selbstverständlich such- und anzeigbar sein. Auch hier galt ursprünglich die oben erwähnte Restriktion hinsichtlich des Weiternavigierens.
- Die ID-Nummer des jeweiligen Normdatensatzes (Unterfeld \$9) wird zwar nicht für die sachliche Facette gebraucht, wäre aber, schon im Hinblick auf die Suchmöglichkeiten, ebenfalls mitzuführen.
- Die Forms Schlagwörter könnten eventuell mit bereits bestehenden formalen Kriterien bezüglich Dokumenttyp bzw. Erscheinungsform kombiniert werden.

Beispiele hiezu zeigt *Abbildung 4*. Aus der ersten Kette (902) wäre zu extrahieren:

¹² Vgl.: Oberhauser, O.; Labner, J. (2006). Entwurf und Implementierung einer neuen Multi-Klassifikations-Normdatei für den Österreichischen Verbundkatalog. B.I.T. online. 9 (4). 293–302.

¹³ Vgl.: Oberhauser, O.; (2008). Sachliche Erschließung im Österreichischen Verbundkatalog: Status und Perspektiven. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare*. 61 (3). 59–77.

¹⁴ Imhof, op. cit.

■ für die sachliche Facette:

1. Wien / Stephansdom
2. Architektur

nicht aber: Führer (da Formschlagwort [Unterfeld \$f])

■ für die Suche:

1. Wien / Stephansdom / Architektur / Führer [Kette]
2. Wien / Stephansdom [Phrase]
2. Wien
2. Stephansdom
2. 4079282-1
3. Architektur
3. 4002851-3
4. Führer

■ für das Display:

1. Wien / Stephansdom; Architektur; Führer [Kette mit Unterscheidung zwischen einfachen und komplexen Schlagwörtern]¹⁵

■ für das Weiternavigieren aus der Vollanzeige:

1. Wien / Stephansdom
2. Architektur
3. Führer [ursprünglich nicht intendiert]

Aus der zweiten Kette (902) wäre zu extrahieren:

■ für die sachliche Facette:

1. Wien / Stephansdom
- nicht aber:** Geschichte (da Zeitschlagwort [Unterfeld \$z])

■ für die Suche:

1. Wien / Stephansdom / Geschichte [Kette]
2. Wien / Stephansdom [Phrase]
2. Wien
2. Stephansdom
2. 4079282-1
3. Geschichte

■ für das Display:

Wien / Stephansdom; Geschichte [Kette mit Unterscheidung zwischen einfachen und komplexen Schlagwörtern]

■ für das Weiternavigieren aus der Vollanzeige:

1. Wien / Stephansdom
2. Geschichte [ursprünglich nicht intendiert]

Aus der dritten Kette (902) wäre zu extrahieren:

■ für die sachliche Facette:

1. Wien / Stephansdom
- nicht aber:** Geschichte 1277-1466 (da Zeitschlagwort [Unterfeld \$z])
- nicht aber:** Quelle (da Formschlagwort [Unterfeld \$f])

■ für die Suche:

1. Wien / Stephansdom / Geschichte 1277-1466 / Quelle [Kette]
2. Wien / Stephansdom [Phrase]
2. Wien
2. Stephansdom
2. 4079282-1
3. Geschichte 1277-1466
4. Quelle

■ für das Display:

1. Wien / Stephansdom; Geschichte 1277-1466; Quelle [Kette mit Unterscheidung zwischen einfachen und komplexen Schlagwörtern]

■ für das Weiternavigieren aus der Vollanzeige:

1. Wien / Stephansdom
2. Geschichte 1277-1466 [ursprünglich nicht intendiert]
3. Quelle [ursprünglich nicht intendiert]

Ein zweites Beispiel zeigt Abbildung 5. Hier soll extrahiert werden:

■ für die sachliche Facette:

1. Hund
 2. Tierernährung
- nicht aber:** Kongress (da Formschlagwort [Unterfeld \$f])
- nicht aber:** Cambridge <1985> (da Geographikum ohne Unterfeld \$9)

■ für die Suche:

1. Hund / Tierernährung / Kongress / Cambridge <1985> [Kette]
2. Hund
2. 4026181-5
3. Tierernährung
3. 4060095-6
4. Kongress
5. Cambridge <1985>

■ für das Display:

1. Hund; Tierernährung; Kongress; Cambridge <1985>

■ für das Weiternavigieren aus der Vollanzeige:

1. Hund
2. Tierernährung
3. Kongress [ursprünglich nicht intendiert]

SW	1.SW-Kette	902	—	s	Hund
SW	1.SW-Kette	902	—	g	4026181-5
SW	1.SW-Kette	902	—	g	Tierernährung
SW	1.SW-Kette	902	—	g	4060095-6
SW	1.SW-Kette	902	—	f	Kongress
SW	1.SW-Kette	902	—	g	Cambridge <1985>

Abbildung 5 und Abbildung 6

Hauptschlagwort	800	—	s	Landschaftsgebundene
Quelle	808	a	a	SWL
s.Verw./Bemerk	830	—	s	Regionale Bauform
s.Verw./Bemerk	830	—	s	Landschaftsgerechtes
ÜbergeordSW	850	—	s	Architektur
ÜbergeordSW	850	—	s	Bauform
VerwandtesSW	860	—	s	Bauernhaus
VerwandtesSW	860	—	s	Volksarchitektur

4. Cambridge <1985> [ursprünglich nicht intendiert]

Extraktion von Normdatenelementen

Wie bereits erwähnt, verfügte bereits der bisherige Verbundkatalog über die Möglichkeit, die katalogisierten Ansetzungsformen mit definierten Datenelementen aus den verwendeten Normdateien (PND, GKD, SWD) anzureichern. Dabei handelte es sich primär um eine Funktionalität zur Erweiterung der Indizes (Wortindex, Phrasenindex) – daher beschränkte sich diese Anreicherung auf die Siehe-Verweise. Im Falle der SWD kamen dafür also die Kategorien 830 (Äquivalente Bezeichnung = Synonym) und 820 (Alternativform zum Hauptschlagwort) in Frage.

Künftig können auch die übrigen Relationierungen – Mehrgliedriger Oberbegriff (845), Übergeordnetes Schlagwort (850) und Verwandtes Schlagwort (860) – miteinbezogen werden. Dies kann etwa am Beispiel des Schlagworts „Landschaftsgebundenes Bauen“ veranschaulicht werden (Abbildung 6).

Für die Vollanzeige bzw. das Weiternavigieren aus dem Display spielt diese Anreicherung keine Rolle, sehr wohl dagegen ist sie für die Suche sowie auch die sachliche Facette von Relevanz. Dafür würde im obigen Beispiel die Extraktion folgende Elemente liefern:

1. Regionale Bauform
2. Landschaftsgerechtes Bauen
3. Architektur
4. Bauform
5. Bauernhaus
6. Volksarchitektur

Nicht vergessen werden soll dabei, dass bestimmte Unterfelder wie z.B. Unterfeld \$v (Bemerkungen), deren Inhalte nicht zur Vokabularerweiterung beitragen, nicht extrahiert werden dürfen.

Problemkreis

Homonymenzusatz

Ein Aspekt, der sowohl die Datenelemente aus den RSWK-Ketten als auch jene aus den Normdaten tangiert, ist der Problemkreis „Homonymenzusatz“ (HZ), der im bisherigen Verbundkatalog nicht zufriedenstellend gelöst werden konnte. Ein künftiges – vielleicht nicht in jedem Detail realistisch – Szenario könnte wie folgt aussehen:

- in der sachlichen Facette soll der HZ angezeigt wer-

¹⁵ Die Verwendung des Semikolons als Trennzeichen zwischen den Schlagwörtern einer Kette ist in Primo aus programmtechnischen Gründen notwendig (Ansetzungsketten behalten dagegen den Schrägstrich).

den, da damit die Unterscheidbarkeit gewährleistet wird

- ebenso soll der HZ in der Vollanzeige sichtbar sein
daraus folgt, dass der HZ auch beim Weiternavigieren mitgeführt wird
- für die einfache Suche soll dies ebenso gelten
- bei der **erweiterten Suche im Schlagwortindex** wäre folgendes Systemverhalten wünschenswert:
 - Stichwörter aus dem HZ sollten weder für sich allein noch in Kombination mit beliebigen anderen Suchbegriffen zu Treffern führen
 - hingegen sollten sie in Kombination mit dem Schlagwort, aus dessen HZ sie ursprünglich stammen, sehr wohl zu Treffern führen
 - nicht betroffen davon sind die gleich-

lautende Suchbegriffe, wenn sie nicht aus dem HZ stammen (also z.B. als Sachschlagwort verwendet werden)

Ein Beispiel wäre etwa das Schlagwort „Holbein, Hans <Künstler, 1497-1543>„: Eine Suche nach „Künstler“ sollte Treffer mit diesem Schlagwort nicht erbringen, eine Suche nach „Holbein Künstler“ hingegen schon. Auch die nachfolgenden, willkürlich ausgewählten Beispiele veranschaulichen diese Problematik:

800 Kirchberg <Regen>
830 Kirchberg <Wald, Bayern>
800 Bank Austria <Schweiz>
800 Oldenburg <Oldenburg, Bezirk>
830 Oldenburg <Land>
830 Oldenburg <Staat>
800 Holbein, Hans <Künstler, 1497-1543>
830 Holbein, Hans <der Jüngere>
800 Ada <Holland, Gräfin>

800 Hunde <Familie>
800 Ölbaum <Motiv>

Kommunikations- und Parametrisierungsgrundlage

Viele der bisherigen Überlegungen sind in der tabellarischen Übersicht in *Abbildung 7*, die auch als Kommunikations- und Parametrisierungsgrundlage für das Primo-Setup gedient hat, zusammengefasst dargestellt. Die Spalten 2–7 beziehen sich auf die unterschiedlichen Sektionen im Primo-Internformat (PNX) bzw. die damit verbundenen Funktionen (Spalte 2 = Vollanzeige; Sp. 3 = Weiternavigieren aus der Vollanzeige; Sp. 4 = feldübergreifende Suche / „Alle-Felder-Index“; Sp. 5 = erweiterte Suche / Sachrecherche; Sp. 6 = Facette „Thema“ (topic); Sp. 7 = Kandidat für weitere Facetten wie Medium oder Form).

1	2	3	4	5	6	7
Extraktion aus	Display display	Display navi	Index search	Index search/subject	Facette Topic	Facette weitere
Titeldaten						
902ff. \$spckgta9	Y	Y	Y	Y HZ=N	Y	N
902ff. \$f	Y	N	Y	N	N	Y
902ff. \$z	Y	N	Y	N	N	Y
902ff. \$g (ohne \$9)	Y	N	Y	N	N	N
Normdaten						
820ff. \$spckgta -\$v	N	–	Y	Y HZ=N	N	N
830 \$spckgt -\$v	N	–	Y	Y HZ=N	N	N
845 \$spckgt -\$v	N	–	Y	Y HZ=N	N	N
850 \$spckgt -\$v	N	–	Y	Y HZ=N	N	N
860 \$spckgt -\$v	N	–	Y	Y HZ=N	N	N

Abbildung 7

In *Abbildung 7* symbolisiert die Farbe Grün die spätere wunschgemäße Implementierung der betreffenden Funktionalität, während die Farbe Rot ausdrückt, dass der ursprüngliche Wunsch (vorerst) nicht realisiert werden konnte. Das bedeutet z.B., dass in Sp. 3 der ursprüngliche Wunsch „N“ bei 902ff. \$f (Form-SW *nicht* für das Weiternavigieren anbieten) *nicht* umgesetzt wurde. Ausgangspunkt sind die Kategorien 902ff.¹⁶ der Titeldaten (Spalte 1 Bereich „Titeldaten“). Wenn bei den Unterfeldern \$spckgta (Sach-SW, Personen-SW, Körperschafts-SW, Geo-SW, Werktitel-SW, Unterschlagwort) über die SWD-Identnummer (IDNR) in \$9 ein Normdatensatz gefunden wird, so werden zusätzlich bestimmte Normdaten-Kategorien (Spalte 1 Bereich „Normdaten“) mit eingebunden. Im anderen Fall (d.h. wenn kein Unterfeld \$9 vorhanden ist bzw. der Inhalt von \$9 zu keinem Normdatensatz führt), so werden nur die Datenelemente aus den Titeldaten herangezogen.

Die Inhalte der Unterfelder \$fz (Form-SW und Zeit-SW) und \$g ohne \$9 (Geo-SW ohne IDNR, z.B. Kongressorte) sollten zumindest für den Schlagwortindex und die Facette „Thema“ unberücksichtigt bleiben. Dies konnte bei der Implementierung nur für die Facette (grünes N in Sp. 6), nicht aber für den Schlagwortindex realisiert werden (rotes N in Sp. 5). Da die genannten Inhalte aber jedenfalls im „Alle-Felder-Index“ vorkommen sollen, konnte die ursprünglich gewünschte Deaktivierung für das Weiternavigieren aus der Vollanzeige nicht erreicht werden (daher rotes N in Sp. 3). Das oben angeführte wünschenswerte Szenario im Zusammenhang mit dem Homonymenzusatz konnte erwartungsgemäß ebenfalls nicht umgesetzt werden (rotes HZ=N in Sp. 5).

Die Möglichkeit der Einbeziehung der synonymen Begriffe bzw. alternativen Ansetzungsformen in der Facette „Thema“ wurde zwar erwogen, schließlich aber doch verworfen (vgl. grünes N in Spalte 6), schon um die Facette nicht unnötig aufzublähen.

In Kauf genommen werden mußte dabei allerdings, dass in Österreich bevorzugte Synonyme (z.B. „Traktor“ anstelle von „Schlepper“) oder auch ins Deutsche übersetzte Werktitel „außen vor“ bleiben müssen. Ob künftig eventuell auch Oberbegriffe (845, 850) in die Facette „Thema“ einfließen sollten, wird erst in Zukunft zu diskutieren sein.

4.2 Andere verbale Sacherschließungselemente

Auch die Einbeziehung der weiteren oben dargestellten verbalen Sacherschließungselemente in das künftige sachliche Such-Interface wurde erwogen. Schließlich kamen wir zu dem Schluss, dass die Recherche nach diesen Datenelementen über den Index „Alle Felder“ realisiert werden soll, nicht aber über den in der erweiterten Suche angebotenen Schlagwortindex (keine Vermischung mit RSWK). Ebenso sollen etwa LCSH *nicht* in der Facette „Thema“ aufscheinen. Sehr wohl hingegen ist es sinnvoll, diese Sacherschließungselemen-

16 d.h. 902, 907, 912, ..., 947.

te in der Vollanzeige darzustellen und auch mit ihnen ein Weiternavigieren zu ermöglichen. Ob die verbalen Elemente aus der BK (Klassenbenennungen, Synonyme) überhaupt für die Suche herangezogen werden sollen oder sogar *nicht* in den Index „Alle Felder“ mit eingehen sollen, wurde ebenfalls diskutiert. Da aber ein Prinzip des Verbundkatalogs darin besteht, immer alle verbalen Datenelemente in diesen Index aufzunehmen, wurde entschieden, den Status Quo auch in der neuen Umgebung beizubehalten. Auch die verbalen Elemente aus der MSC sollen in diesen Index aufgenommen werden.

Primo hat, ganz dem Anforderungsprofil entsprechend, eine weitere, sehr wichtige Verbesserung mit sich gebracht, nämlich die Möglichkeit der völligen Integration der zu den Katalogisaten angelegten Abstracts, Inhaltsverzeichnisse und Klappentexte. Alle drei genannten Dokumenttypen sollen im Volltext (d.h. für den Index „Alle Felder“) indiziert werden, wie das auch in eDoc der Fall war. Man kann diese Stichwörter als unkontrolliertes Vokabular mit potentiell mehr oder weniger ausgeprägtem inhaltserschließenden Charakter betrachten.

4.3 Klassifikatorische Sacherschließungselemente

Die im bisherigen Verbundkatalog angebotene Funktion „Index blättern“ steht in Primo nicht mehr zur Verfügung. Dies betrifft auch die fünf oben angeführten Verbundklassifikationen, für die ein Sucheinstieg über den Index sehr nützlich war. Bei der auf dem Stichwortprinzip basierenden Suche sollen einerseits die Notationen bzw. ihre Bestandteile (z.B. im Fall der RVK: PB 1601 bzw. PB und 1601) als Suchbegriffe eingegeben werden können, andererseits – bei den beiden als Normdatei verfügbaren Systemen (BK, MSC) – auch die Stichwörter aus den verbalen Elementen. Neben der Möglichkeit, diese Datenelemente im Index „Alle Felder“ (einfache Suche) zu recherchieren, soll die „erweiterte Suche“ einen feldspezifischen Einstieg pro Klassifikationssystem anbieten.

Für die Darstellung in der Vollanzeige und das Weiternavigieren aus dieser wären die vollständigen Notationen (inklusive der Klassenbenennungen bei BK und MSC) vorzusehen. Mit der im bisherigen OPAC gebotenen Möglichkeit der „Interpretation“ der Notationen der RVK durch die Verlinkung mit der Regensburger Datenbank¹⁷ bzw. der DDC und der ZDB-Systematik mittels hin-

terlegter Listen kann (vorerst) wohl nicht gerechnet werden.

Die Frage, inwieweit eine der Verbundklassifikationen für eine Facette (oder auch mehrere für mehrere Facetten) genutzt werden sollte, ist aufgrund des derzeitigen Anteils der mit Notationen ausgestatteten Datensätze (wie oben erwähnt, hält diesbezüglich die RVK mit 7,5% die Spitze) eher negativ zu beantworten. Sollte es dagegen gelingen, durch verbundübergreifende Anreicherungsprojekte eine „flächendeckendere“ Ausstattung mit Notationen zu erreichen,¹⁸ wären insbesondere die BK, aber auch die RVK oder die DDC für eine Facette geeignet. Wie aus anderen Primo-Applikationen bekannt ist,¹⁹ bietet Primo für die Library of Congress Classification (LCC) sogar eine hierarchische Facetten-Option an, bei der zunächst nur die oberste Hierarchiestufe angezeigt wird. Bei Auswahl einer Hauptklasse wird die Facette neu berechnet und die nächste darunter liegende Hierarchiestufe zur Auswahl angezeigt. So kann ein Ergebnisset immer enger eingeschränkt werden, ohne sofort auf die kleinste Hierarchiestufe wechseln zu müssen. Letztere wäre aufgrund meist geringer Fallzahlen – ähnlich wie im Fall ganzer RSWK-Ketten – ohnedies für eine Facette weniger geeignet.

5 Implementierung

Wie bereits in Abschnitt 2 kurz angerissen, wird jeder Datensatz in Primo als XML-Datei abgespeichert. Die einzelnen Sektionen dieser XML-Datei werden für jeweils unterschiedliche Funktionen von Primo ausgewertet. Über die Normalisierung ist jede MAB-Kategorie, jeder Indikator, jedes Unterfeld getrennt ansprechbar und kann einer oder mehreren PNX-Abschnitten zugeordnet werden. Einzelne MAB-Felder und deren Unterfelder können zusammengeführt oder zerschlagen werden, zahlreiche String-basierte Funktionen können das Katalogisat anreichern oder beschneiden. Die Trennung von Erfassungs- und Präsentationssystem (Aleph und Primo), die übersichtliche Aufteilung der XML-Abschnitte und die einfache Handhabung der Transformation ermöglichen somit eine gezielte Manipulation der Quelldateien, einen maßgeschneiderten Zuschnitt für die spätere Verwendung und in der Folge eine nachträgliche Bereinigung der Datenbank. Es ist evident, dass ein sauberes Quellsystem

erste Wahl bleibt – Primo sollte als Datenempfänger, d.h. abhängiges System, nicht zur Kaschierung fehlerhafter Quellen missbraucht werden. Dennoch: Eine Bereinigung von Altlasten oder fehlerhaften Katalogisaten scheitert oft an Ressourcenknappheit; für solche Fälle ergibt sich mit Hilfe der Normalisierung die Möglichkeit, mehr oder weniger umfassende Katalogbereinigungen durchzuführen.

Noch vor der Normalisierung muss ein Weg gefunden werden, die Daten in einem für Primo lesbaren Format aufzubereiten. In Aleph steht für diese Aufgabe ein eigenes Tool, der *Aleph Publishing Mechanism* (APM), zur Verfügung. Dieser erlaubt es, an bestimmten „Stellschrauben zu drehen“ und damit den Aufwand auf der Seite von Primo bereits um einiges zu reduzieren. Die bestehende bibliographische Datenbank wird quasi im XML-Format dupliziert. Jeder einzelne Datensatz kann zusätzlich um Angaben aus übergeordneten Werken, Bestandsangaben für Exemplare und Zeitschriften sowie Normdaten-Verlinkungen angereichert werden, was gleichsam zu einer Aufhebung der relationalen Struktur und zu einer Vervielfachung der katalogisierten Felder führt. Ein abhängiger Band z.B. ist dann nicht mehr über einen Link mit der Gesamttitelaufnahme verbunden, sondern enthält alle Angaben des übergeordneten Werkes. Über den APM wurde auch die oben angesprochene Anreicherung der Sätze mit über- und gleichgeordneten Schlagwörtern realisiert.

Im Rahmen dieses Beitrags sind insbesondere drei PNX-Sektionen von Interesse: *display*, *search* und *facet*. Entsprechend der Tabelle in *Abbildung 7* wurden die in den Feldern 902ff. abgelegten RSWK-Ketten für die Facette „Thema“ zerschlagen und jene Felder, bei denen es sich um Geographika oder Forms Schlagwörter handelt, entfernt. Dem besonderen Charakter einer RSWK-Kette, der syntaktischen Indexierung, wird durch die Zerschlagung übrigens einmal mehr *nicht* Rechnung getragen. Von der angesprochenen Möglichkeit der Katalogbereinigung wird auch in der thematischen Facette Gebrauch gemacht; so werden z.B. die unterschiedlichen Schreibweisen von Forms Schlagwörtern („Kongress“ und „Kongreß“) zusammengeführt.

Was die Titelvollanzeige betrifft, so wurde für diese von einer Zerschlagung der Ketten abgesehen. Deren Anzeige erfolgt wie gewohnt: eine Kette pro Zeile. In diesem Zusammenhang sei auf eine Standardfunktion von Primo hingewiesen, die aus bestimmten Feldern der Vollanzeige automatisch einen Hyperlink generiert, der dann für eine weiterführende Suche angeklickt werden kann (an jenen Stellen, an denen bereits von

17 Vgl.: Oberhauser, O. (2009). Das RVK-Interface der FU Berlin – modifiziert, zentral gehostet und lokal im Online-Katalog der WU Wien genutzt. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*. 62(3). 43–56.

18 In diesem Fall wäre auch der Bedarf für ein hierarchisches Such-Interface à la *MelvilleSearch* für die DDC (<http://melvil.d-nb.de/>) oder *RVK-Online* (<http://rvk.uni-regensburg.de/>) gegeben.

19 So z.B. in den Katalogen der Universitäten Vanderbilt (<http://discoverlibrary.vanderbilt.edu/>), Minnesota (<http://www.lib.umn.edu/catalogs>) oder Guelph (<http://www.lib.uoguelph.ca/>).

„Weiter navigieren“ die Rede war, war genau diese Funktion gemeint). Dies gilt zunächst für das Feld „Autor/ in“, aber auch für die einzelnen Elemente einer Schlagwortkette: Jedes Element einer Kette kann für eine weiterführende Suche ausgewählt werden. Wäre die gesamte Kette ein einziger Link und würde ein Klick darauf eine Suche nach Werken mit exakt dieser RSWK-Kette durchführen, dann wäre wegen der eher seltenen mehrfachen Verwendung ein und derselben Kette der Sucherfolg oft ziemlich bescheiden (nämlich ein einziger Treffer bzw. nur sehr wenige Treffer). Mit der Zersplitterung der Ketten in ihre Bestandteile für die Bildung der Ausprägungen in den Facetten bzw. die weiterführende Suche in der Vollanzeige geht einher, dass auch die Indexeinträge nicht aus ganzen Ketten, sondern aus einzelnen Schlagwörtern bestehen; andernfalls würde die weiterführende Suche ins Leere gehen. Unsere Parametrisierung sieht vor, dass im Anschluss an eine *einfache Suche* das Weiter navigieren mit Schlagwörtern im Index „Alle Felder“ erfolgt, im Anschluss an eine *erweiterte Suche* hingegen im Schlagwortindex. Wenn im Index jedoch z.B. der *angezeigte* Homonymenzusatz nicht vorhanden wäre, würde eine weiterführende Suche ins Leere gehen, da etwa der Suchbegriff „Holbein, Hans <Künstler, 1497-1543>“, den

Indexeintrag „Holbein, Hans“ nicht finden würde. Somit ist der ursprünglich gewünschte Ausschluss der Homonymenzusätze aus dem Schlagwortindex nicht realisierbar. Ein entsprechender Wunsch bezüglich einer Einflussnahme auf die Linkbildung wurde bei der Herstellerfirma bereits deponiert.

Bei einer Suchmaschine, die in der Lage ist, eine Gewichtung der gefundenen Treffer vorzunehmen, stellt sich unweigerlich die Frage, wie thematische Aspekte zu gewichten sind. In diesem Zusammenhang tritt der Vorteil kontrollierter Metadaten gegenüber Volltextdokumenten deutlich zu Tage: Da der Suchmaschine eindeutig mitgeteilt werden kann, in welchem Feld welche Information zu finden ist, ist es ohne weiteres möglich, bestimmte Felder bei der Berechnung der Relevanz höher zu gewichten als andere.²⁰ Natürlich findet eine unterschiedliche Gewichtung verschiedener Felder auch bei

Internet-Suchmaschinen statt. Was diese aber teilweise erst mühsam ermitteln müssen, liegt bei Katalogdaten unzweifelhaft vor. Auch handelt es sich um einwandfreie und nicht – wie bei manchen Web-Dokumenten – mit manipulativen Absichten (um ein besseres Ranking zu erzielen) vergebene Indexbegriffe. Im Gegensatz zu Web-Suchmaschinen dürfte die thematische Suche, die Suche, die ein Informationsbedürfnis zu einem bestimmten Wissensgebiet befriedigen soll, in Bibliothekskatalogen neben der Suche von Werken von bestimmten Autoren bzw. nach mehr oder weniger bekannten Werktiteln die häufigste Form der Suche darstellen. Ein kurzer Blick auf die bis dato in Primo eingegebenen Suchstrings bestätigt diese Vermutung. Benutzer suchen zumeist nach Autoren („Simone de Beauvoir“), Werken („Erich Fromm Kunst des Liebens“) oder, am häufigsten, thematisch („Essstörung Kinder“). Demgegenüber zeigt die Auswertung von Suchen in Internet-Suchmaschinen ein viel breiter gefächertes Spektrum an Intentionen, die der Suche zu Grunde liegen.²¹ Folgerichtig erhalten

20 Vgl. z.B.: Dellit, A.; Boston, T. (2007). Relevance ranking of results from MARC-based catalogues: From guidelines to implementation exploiting structured metadata. Paper presented at *Information online 2007*, Sydney, NSW; 30th January – 1st February 2007.

Online: <http://conferences.alia.org.au/online2007/Presentations/30Jan.B1.relevance.ranking.pdf>;

Langenstein, A.; Maylein, L. (2009). Relevanz-Ranking im OPAC der Universitätsbibliothek Heidelberg. *B.I.T. online*. 12 (4). 408–413; Lewandowski, D. (2009). Ranking library materials. *Library hi tech*. 27 (4). 584–593.

21 Rose, D. E.; Levinson, D. (2004): Understanding user goals in web search. *Proceedings of the 13th International Conference on World Wide Web, WWW '04*. New York: ACM. 13–19. DOI: <http://doi.acm.org/10.1145/988672.988675>

alle RSKW-Schlagwörter einen Multiplikator, der sie vor gleichlautenden Begriffen in anderen Abschnitten des Datensatzes auszeichnet. Dies ist allerdings eine Voreinstellung des Herstellers, in die bisher von unserer Seite nicht eingegriffen wurde.

Was die weiteren verbalen Erschließungselemente betrifft, so wurde während der Phase der technischen Implementierung ganz besonderes Augenmerk auf die Indexierung von Abstracts, Inhaltsverzeichnissen sowie Klappentexten gelegt. Wie sich herausstellte, waren Planung und Umsetzung schwieriger und erforderten mehr Zeitaufwand als vorher angenommen. Dennoch konnten schließlich die textuellen Anteile aller 300.000 betroffenen, im System eDoc abgelegten Objekte extrahiert und in Primo geladen werden. Dies wurde über ein sogenanntes Plug-in – Primo bietet die Möglichkeit, die fertigen PNX-Sätze noch weiter mit eigenen Java Programmen zu bearbeiten – realisiert. Anfängliche Befürchtungen, die Volltextindexate könnten eine „Verwässerung“ der Metadaten mit sich bringen, haben sich bislang nicht bestätigt. Sollte jedoch Bedarf bestehen, so kann die Gewichtung von Volltextindexaten beeinflusst werden. Letztlich ist deren Menge jedoch noch zu gering, um einen ernsthaften Einfluss auf Suchergebnisse auszuüben.²² Darüber hinaus steht es jeder teilnehmenden Einrichtung frei, ob in ihrer Sicht eine Volltextindexierung überhaupt zum Einsatz kommen soll oder nicht. Die Normalisierung der Klassifikationskategorien erwies sich im Grunde als kaum problembehaftet, da der Kategorieninhalt – d.h. die Notation – meist vollständig kopiert wird. Leider stehen, wie oben erwähnt, Klassenbenennungen und andere verbale Elemente nur für die BK und die MSC (für die eine Normdatei vorgehalten wird), zur Verfügung. So wird etwa die in Aleph erfasste BK-Notation über den APM übernommen, aus der Normdatei mit Klassenbenennung und Synonymen ergänzt, wodurch Notation und verbale Bezeichnung angezeigt und facettiert werden können. Dass die BK wegen ihrer zu geringen Verbreitung derzeit noch nicht als Facette angezeigt werden kann, wurde bereits weiter oben vermerkt. Dagegen werden im Fall der RVK in Aleph nur die Notationen katalogisiert, was für eine Anzeige bzw. Facette wenig benutzerfreundlich ist. Ex Libris bietet im Auslieferungsumfang von Primo für dieses Problem ein Anreicherungsverfahren an, welches eine bestimmte RVK-Notation in eine verbale Bezeichnung übersetzt, die als die jewei-

lige Ausprägung der RVK-Facette angezeigt wird. Von diesem Verfahren wird aber bei uns kein Gebrauch gemacht, da diese Übersetzung auf dem Hinzufügen der Klassenbenennungen aus der gesamten Hierarchie beruht und so zu unübersichtlichen und viel zu langen Ausprägungsbezeichnungen führt (so etwa für die Notation ZB 86161: „Land- und Forstwirtschaft. Gartenbau. Fischereiwirtschaft. Hauswirtschaft – Wirtschafts- und Sozialwissenschaften – Agrarmarktpolitik – Interventionsinstrumente – Asien, Orient, Östliche Welt – Zentral und Ostasien, Hinterer Orient – China (einschl. Tibet, Mandschurei, Taiwan, Hongkong, Macao)“). Ideal wäre dagegen eine Form der hierarchischen Facettierung wie bereits oben für die LCC beschrieben. Als Kompromiss zwischen der allerengsten, über das Anreicherungsverfahren ermittelten Klassenbenennung und gar keiner Benennung findet daher die Facettenbildung derzeit auf zweiter Hierarchieebene statt. ZB 86161 fällt somit in die Ausprägung „ZB: Agrarwissenschaftliche Ausbildung, Agrarwissenschaftliche Forschung, Agrarverwaltung und -recht sowie Wirtschafts- und Sozialwissenschaften“. Die Umsetzung Notation à Benennung findet ausschließlich in der Facette statt – eine verbale Suche nach Klassenbezeichnungen ist weder in der einfachen, noch in der erweiterten Suche möglich. Inwieweit hier noch eine Verbesserung (d.h. Suchbarkeit der verbalen Elemente) erzielt werden kann, müsste erst eruiert werden. Aus bereits erläuterten Gründen muss der Index wiederum der Anzeige entsprechen – andernfalls würde ein Weiternavigieren keine Resultate erbringen. Was die Gewichtung der Notationsfelder betrifft, besteht auch hier die Möglichkeit, steuernd einzugreifen. Bisher hat sich dafür aber keine Notwendigkeit ergeben. Die abschließliche Suche mit Notationen (d.h. alphanumerischen Codes) ist im Benutzerverhalten so gut wie überhaupt nicht registrierbar, was auch das Hervorheben oder Herabstufen der entsprechenden Felder unnötig erscheinen lässt. Sollte (dereinst) die erwähnte Suchbarkeit der Klassenbenennungen der RVK erreicht werden, könnte eine Neubewertung dieses Aspekts erforderlich sein.

Die grundsätzliche Problematik der Facettierung bei Beständen, in denen nicht jeder Datensatz jeder Facette zugeordnet werden kann, wurde bereits gestreift (vgl. Abschnitt 4.1, erster Absatz). Sie liegt darin, dass durch Auswahl einer Facetten-Ausprägung nicht nur die Datensätze ausgeschlossen werden, die *einer anderen* Ausprägung zugeordnet sind, sondern auch jene, die *gar keine* Ausprägung besitzen. Aus diesem Grund wird bei der Bildung der Facetten für jeden Satz zumindest ein Dummy-Wert vergeben. So

enthält jede Facette auch eine Ausprägung „Ohne Angabe“ oder „Nicht klassifiziert“, welches meist jenes ist, das die höchste Trefferzahl aufweist. Ob dadurch tatsächlich das Bewusstsein dafür gestärkt werden kann, dass die Auswahl einer Facette immer mit dem Verlust einer großen Anzahl dort nicht verteilter Datensätze, die gleichwohl relevant gewesen sein könnten, mit sich bringt, sei allerdings dahingestellt.

6 Ausblick

Mit der Einführung von Primo hat eine neue Ära der wissenschaftlichen Recherche im Österreichischen Bibliothekenverbund begonnen. Suchmaschinentechnologie (Ergebnissatz gewichtet nach Relevanz, facettierte Navigation) sowie die Vereinigung von Metadaten und Volltexten sind die Hauptcharakteristika der Innovation. Die daraus folgende Umstrukturierung und Neuordnung der sacherschließenden Elemente im Index und in der Suchoberfläche war nicht unbedeutend und erlaubt eine noch gezieltere Recherche nach inhaltlichen Aspekten.

Die Vorzüge der neuen Lösung können aber eigentlich erst voll zum Tragen kommen, wenn das vorhandene Datenmaterial bestimmten Qualitäts- und Quantitätsansprüchen genügt. War etwa die lediglich partielle Ausstattung der Datensätze mit RSWK-Ketten auch schon bisher ein Problem – die Schlagwortrecherche reduzierte, meist unbemerkt, die Suchbasis um mehr als die Hälfte –, so gilt dies umso mehr für die Facette „Thema“ (bei der sich regelmäßig die Ausprägung „Ohne Angabe“ als erste positioniert), zumal „gute“ Facetten – wie oben bereits zitiert – im Idealfall auf Kriterien basieren sollten, die bei allen Datensätzen vorkommen. Das bedeutet, dass zukünftige Aktivitäten in vermehrtem Masse darauf abzielen müssen, die Verbunddatensätze durch Kataloganreicherungen (in verbundübergreifenden Kooperationsprojekten) mit RSWK-Ketten, aber auch mit Elementen der klassifikatorischen Sacherschließung möglichst flächendeckend auszustatten. Notationen der BK, RVK und DDC wären hier besonders hervorzuheben, aber auch verbale Elemente (Klassenbenennungen, Verweisungsformen usw.), die bei RVK und DDC zur Zeit noch nicht über Normdateien oder andere Quellen in zufriedenstellender Weise zur Verfügung stehen. Zu den Sacherschließungselementen zählen im weiteren Sinn auch die Stichwörter aus den elektronischen Dokumenten (Abstracts, Klappentexte, Rezensionen, Inhaltsverzeichnisse etc.) Die Forderung nach einer verstärkten Kataloganreicherung gilt daher auch für diese Objekte.

22 Das könnte sich eventuell ändern, wenn die Österreichische Nationalbibliothek den Primo-Betrieb aufnimmt. Vor kurzem wurde bekannt, dass Google einen guten Teil der wertvollen Bestände der Nationalbibliothek einscannen wird (<http://www.onb.ac.at/bibliothek/austrianbooksonline.htm>).

Aus dem oben Gesagten geht hervor, dass bisher nicht alle Intentionen hinsichtlich der inhaltlichen Recherche realisiert werden konnten. Als Defizit kann bspw. vermerkt werden, dass aus bereits genannten Gründen der Schlagwortindex derzeit auch die Form-, Zeit- und Geo-Schlagwörter (ohne IDNR) enthält, sodass eine Recherche nach einem davon betroffenen Begriff, der als Sach-SW gesucht wird, nahezu aussichtslos ist. Eine denkbare Lösungsvariante kann kurz wie folgt skizziert werden: Der bisherige Schlagwortindex, den auch die *display section* benötigt, um die vollständigen RSWK-Ketten in der Vollanzeige darzustellen, würde unverändert bleiben. Daneben könnte aber ein zweiter Sachindex definiert werden, der die unerwünschten Form-, Zeit- und Geo-Schlagwörter *nicht* enthält. In der „erweiterten Suche“ würde dann dieser neue Index anstelle des vollständigen Schlagwortindex angeboten werden. Die oben ebenfalls diskutierten Probleme hinsichtlich der Homonymenzusätze und der nicht unterdrückbaren Weiternavigation mit Form-, Zeit und Geo-Schlagwörtern lassen sich durch eine solche Lösung allerdings nicht bereinigen.

Kurz erwähnt wurde bereits, dass für die klassifikationsbasierten Facetten (RVK, künftig ev. auch BK und DDC) ein hierarchischer Navigationsmechanismus nach dem Muster der LCC (in nord-amerikanischen Primo-Implementierungen) wünschenswert wäre. In diesem Zusammenhang werden vermutlich neue Herausforderungen entstehen, da sich RVK und BK aufgrund ihrer nicht bzw. nur teilweise die Hierarchien abbildenden Notationen nicht ohne zusätzlichen Aufwand in ein hierarchisch arbeitendes Interface integrieren lassen.

Primo – ein relativ junges Produkt – wird vom Hersteller laufend weiterentwickelt. Zurzeit ist die neue Version 3 bei der OBVSG in einer Testphase; sie bringt als Neuerungen schwerpunktmäßig lokale OPAC-Funktionalitäten (Ausleihen, Vormerken etc.) Wir sind optimistisch, was auch die Weiterentwicklung auf dem Retrievalsektor betrifft und erwarten Verbesserungen hinsichtlich der inhaltsbezogenen Recherche in naher Zukunft.

Danksagung

Die Autoren danken Johann Brandauer und Martin Hekele (beide OBVSG) für die Durchsicht des Manuskripts.

■ AUTOREN

MAG. MARKUS KNITEL

Systembibliothekar in der OBVSG, betreut die zentrale Primo-Instanz des Österreichischen Bibliothekenverbundes.
markus.knitel@obvsg.at



ING. JOSEF LABNER

Systemanalytiker im Österreichischen Bibliothekenverbund, in den Bereichen Normdateien, Datenformate und Systemparametrisierung.
josef.labner@obvsg.at



DR. OTTO OBERHAUSER MPhil MSc MLIS

Im Österreichischen Bibliothekenverbund vor allem in den Bereichen Online-Kataloge (zentrale und lokale Systeme) und Systemparametrisierung tätig.
otto.oberhauser@obvsg.at



Alle:

Die Österreichische Bibliothekenverbund und Service GmbH (OBVSG), Brünllbadgasse 17/2a, A 1090 Wien, Österreich
<http://www.obvsg.at>